1-1-2016

Cohomology Operations On Random Spaces

Matthew John Zabka
Wayne State University,

COHOMOLOGY OPERATIONS ON RANDOM SPACES

by

MATTHEW ZABKA

DISSERTATION

Submitted to the Graduate School,

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2016

MAJOR: MATHEMATICS

Approved By:

Advisor	Date

DEDICATION

To my parents

ACKNOWLEDGEMENTS

It is no exaggeration to say that I never could have finished my doctorate without my advisor Daniel Isaksen. His support and encouragement were indispensable throughout this process. It would not be possible to express enough gratitude to him, and I cannot imagine a better advisor.

I appreciate all of the many colleagues and friends I have had at Wayne State. The faculty and staff of the Wayne State Mathematics Department have obviously contributed greatly to my understanding of mathematics, but have also helped me develop as a mathematics instructor. I am extremely grateful for their guidance.

I should also like to thank my parents, brothers, and Anabel Stöckle for their love and support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Clustering

The speed of modern computers has made it possible to study large data sets in high dimensions. Statisticians are often interested in such data's structure or shape, and there are several traditional methods to investigate these properties. For example, principal component analysis projects high-dimensional data onto a subspace we can see, usually $\mathbb{R}^2$ or $\mathbb{R}^3$, in such a way to maximize the variance of the projected data. Such a projection, however, can lose much of the data's structure. Topological data analysis proposes several different approaches using the tools of topology to capture the data's structure.

One type of topological data analysis examines the structure of a topological space generated by data. A common way to generate a topological space has its roots in a statistical method called clustering. As its name suggests, the goal of clustering is to find clusters of data – points that lie close to each other and are far from data in other clusters – in the data set. The most intuitive clustering algorithm is called single-linkage clustering, which we outline below.

Let X be a set of data in a metric space with metric d, and let $r > 0$. Our goal is to create a partition of X based on r into sets we shall call clusters. For two data x and y in X, we say a chain exists between x and y if and only if there exist some data $x_1, x_2, \ldots x_n$ in X such that $d(x, x_1) < r$, $d(x_n, y) < r$, and $d(x_i, x_{i+1}) < r$ for all i in $\{1, 2, \ldots, x_{n-1}\}$. It is easy to see that the property of x and y being connected by a chain is an equivalence relation. The partition sets defined by this equivalence relation are called clusters. It is important to note that clustering depends on the choice of the parameter r. For a sufficiently small r ($r < \min_{x,y \in X} d(x, y)$) every datum belongs to only the singleton cluster containing only itself. For sufficiently large r ($r \geq \max_{x,y \in X} d(x, y)$) the only cluster is the entire set of data X.

Thus for any $r > 0$, we can partition our data into clusters. In data analysis, we seek to answer the following question: What values of r give us clusters that represent genuine structure in our data? To answer this question, we note that if a cluster C persists over a large range of values for r, then the points of C must be far away from the data that are not in C. Therefore, we can conclude that clusters that persist over a large range of r are genuine to the structure of the data set.

A good way to see which clusters persist over a large range of r is to use a dendrogram. We give an

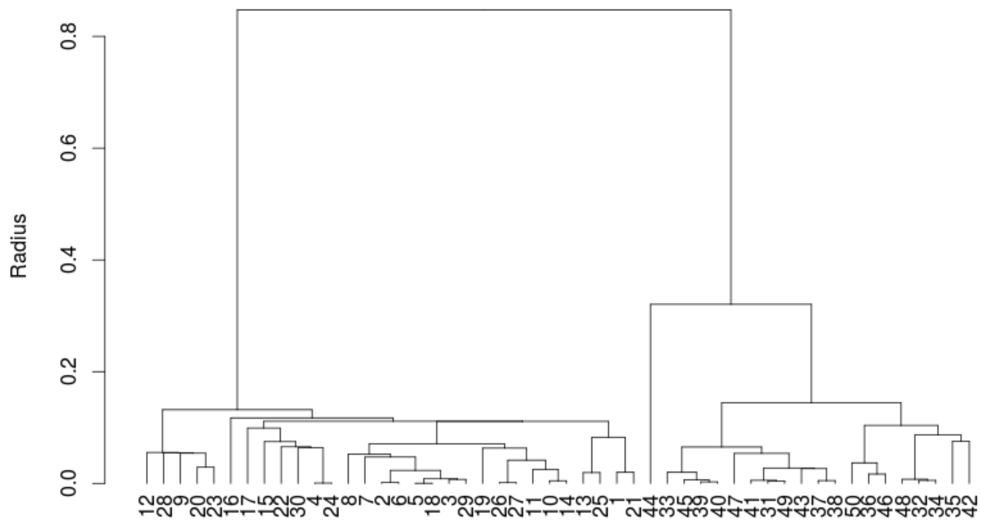


Figure 1: Dendrogram of single linkage clustering on simulated data with two clusters

example in Figure 1 of a dendrogram from fifty simulated points. Each number on the x-axis represents one of the fifty points, while the y-axis represents the size of r. At $r = 0$, each datum is its own cluster. As r increases, we can see how clusters merge together to form new clusters until r is about 0.82, where only one cluster exists. From this picture, we can see that two clusters persist over a large range of r and conclude that these two clusters represent genuine structure of the data. Indeed, these simulated data come from two different normal distributions with distinct means and small variances.

Statisticians have used single-linkage clustering since at least the 1940s. One can consider the single-linkage clustering method topologically as follows: Let X be a data set and $r > 0$. Consider the graph $G(X, r)$, which has vertices X and an edge between two vertices x and y if and only if $d(x, y) < r$. By definition, x and y are members of the same connected component of $G(X, r)$ if and only if x and y belong to the same cluster in the single-linkage clustering when the radius is r. To find structure in the data, we again would see which connected components persist over large ranges of r.

1.2 Persistent Homology

We should like to generalize the notion of finding the connected components of a graph generated by data in order to analyze the data's higher dimensional structure. That is, we should like to generate a topological space from the graph constructed from the data and analyze its topology. To do this, for each positive integer n and for each set of $n+1$ points in $G(X, r)$, we must decide whether to include the n-simplex consisting of these $n+1$ points in our space.

Given $n+1$ points $x_0, x_1, \cdots, x_n$ in our data set X, there are two common methods for determining whether an n-simplex connects these points. The Čech complex is the space with vertices consisting of our data X and the n simplex connecting the data points $x_0, x_1, \cdots, x_n$ whenever $\cap_{i=0}^{n} B(x_i, r/2) \neq \emptyset$. The Vietoris-Rips complex on the other hand contains this n-simplex whenever $B(x_i, r) \cap B(x_j, r) \neq 0$ for all i and j between 0 and n. For a given r and data set X, it follows immediately from the definitions that the Čech complex is a subcomplex of the Vietoris-Rips complex.

Using either of these methods, one may construct a topological space from data. We should like to analyze the structure of these spaces using the tools of topology. One method, known as persistent homology, generalizes clustering by looking at the homology or cohomology groups of a complex. We ask which generators of the homology groups persist over different ranges of r and conclude that, if a generator indeed persists, it tells us something about the structure of our data. Note that the rank of the n-th homology group of a space is called the n-th Betti number. The 0-th Betti number counts the number of connected components of a space. Again, if the space is generated by data, for a given $r > 0$, the connected components are exactly the clusters, and thus persistent homology generalizes single-linkage clustering.

In persistent homology, the parameter r is often called the time parameter. We again should like a way to visualize which generators of homology persist over large values of time. One way to do this is with a barcode diagram. In Figure 2, we have provided the barcode for simulated data taken uniformly on the wedge of two 1-spheres. The x-axis represents the radius r. Each red bar represents a generator of H_1, while each black bar represents a generator of H_0, that is, a connected component. We see that one generator of H_0 and two generators of H_1 persist over a large range of r, which is exactly what we should expect for data taken on $S^1 \vee S^1$.

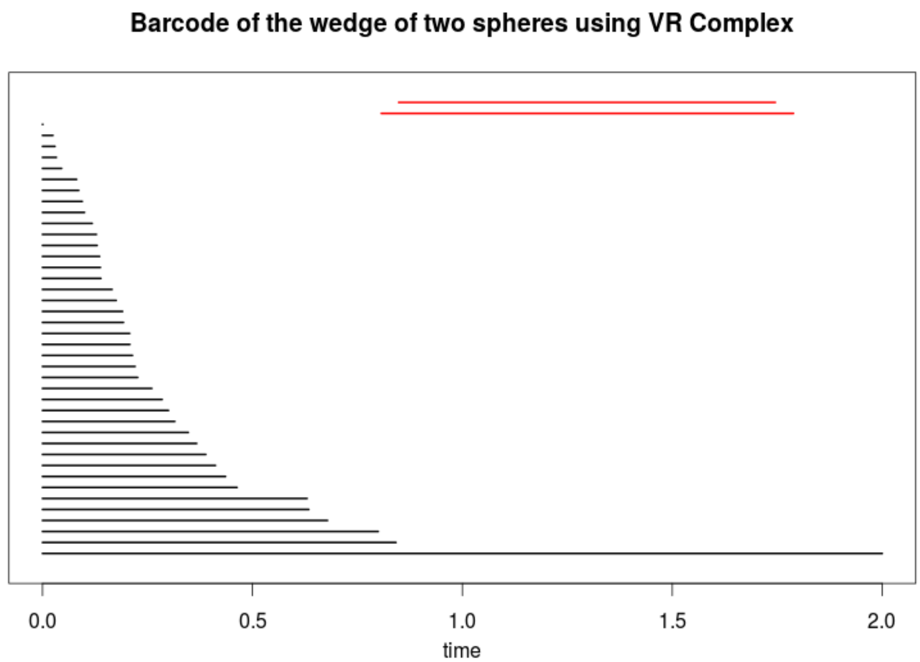


Figure 2: Barcode of data taken uniformly on $S^1 \vee S^2$

If there are many points in a data set, it could be difficult to interpret a barcode. In Figure 3, we simulate five-hundred points on a torus and create the barcode. With so many bars, it is difficult to see when they are born and when they die.

Thankfully, there are other ways to visualize the persistent homology of a space. Each generator of the homology a space has a time at which it is born and a different time at which it dies – except for a single generator of H_0, which never dies. This can be easily seen on our barcode in Figure 2. Note that both generators of H_1 are born at approximately time $r = 0.8$ and die at time $r = 1.75$.

Thus, every bar has a starting point and an ending point – a birth time and a death time. So every bar in a barcode can be thought of as a point in $\mathbb{R}^2$ whose first coordinate is its birth time and second coordinate is its death time. We can then plot these points. Since the birth time must be earlier than the death time, all points will lie above the line $y = x$. Generators that lie farthest from this line persist the longest and thus represent structure in the data. These scatter plots are known as persistence diagrams. We provide the persistence diagram for the data on the torus in Figure 4.

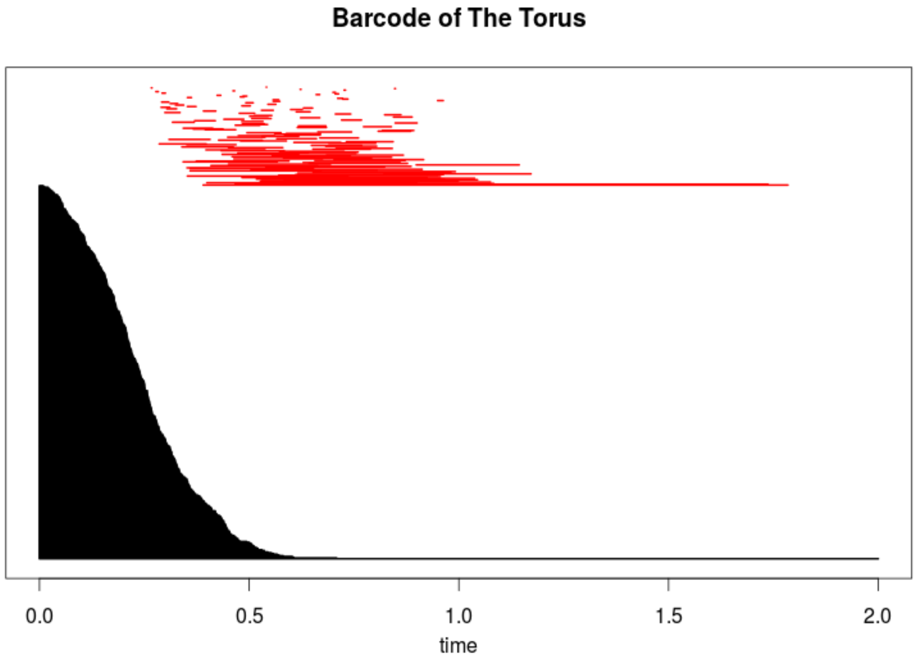


Figure 3: Barcode of 500 points taken uniformly on the torus

The black circles represent generators of H_0, while the red triangles represent generators of H_1. We see one black dot and two red triangles far away from the line $x = y$. This means that one generator of H_0 and two generators of H_1 persist over a large range of values of r, which is exactly what we should expect for data on the torus.

Note that one could use any clustering method to create the one-skeleton (the graph) of our topological space. Since the Vietoris-rips complex depends only on the one-skeleton, one could still generate a topological space if one used a different clustering method. For example, complete clustering stipulates that x be in the cluster C if and only if $d(x, y) < r$ for every y in C.

Persistence homology has proved to be quite useful. For example in [**?**], Lee, Pedersen, and Mumford used persistent homology to analyze the structure of pixels of natural images. In an image, a number is assigned to every pixel. This number is known as the gray scale. Thus, if P is the number of pixels in an image, the image may be thought of as living in $\mathbb{R}^P$. We could then ask about the nature of the collection of natural images – meaning images of actual things as opposed to random noise – lying in $\mathbb{R}^P$. For example,

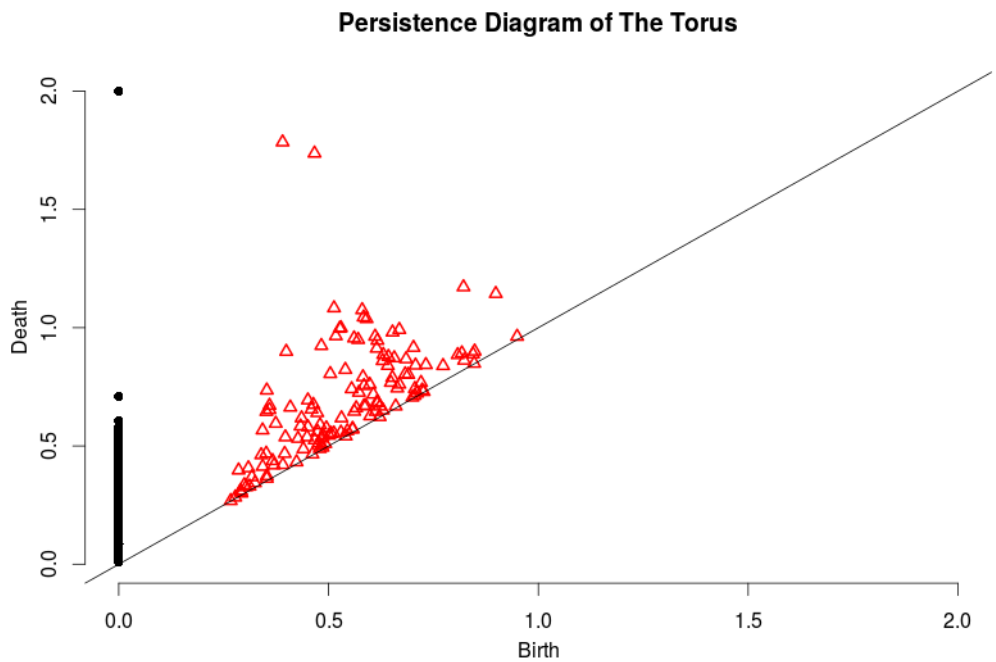


Figure 4: Persistence Diagram of 500 points taken uniformly on the torus

can this collection be modelled as submanifold of $\mathbb{R}^P$?

To answer this question, the authors had to cleverly decide which pixels in an image to study. First, the authors considered 9×9 patches of pixels. Since there are many such patches in an image, $5,000$ such patches were randomly selected. Then, using a norm to determine which patches contained the most variance in gray scale, the authors analyzed the 20% of the 9×9 patches with the most variance.

The reason for only analyzing the patches with the most variance in gray scale is because these patches contain the most information in the data. For example, if one imagines a picture of a landscape with green field and blue sky, then all of the pixels in 9×9 patches contained in only the field or only the sky would have the nearly the same gray scale values and would not be interesting to study. On the other hand, 9×9 patches that contain both the field and the sky will have more interesting structure to analyze.

After a few more tweaks, the authors performed persistent homology on the selected pixels from the natural images and found a space that has the same homology as the Klein bottle. Thus, one has the interesting result that one could use algebraic topology to distinguish natural images from images containing only random noise.

1.3 The Mapper Process

Besides persistent homology, there are also other useful topological tools for data analysis. In [?], Nicolau, Levine, and Carlsson used what is called the mapper process to distinguish a new type of breast cancer. The mapper process was first outlined by Gurjeet Singh, Mémoli, and Carlsson in [?]. This process can be viewed as a way of assigning coordinates to a topological space lying in high-dimensional space in order to visualize it. In particular, the mapper process assigns to a high-dimensional data set a simplicial complex that can be realized in $\mathbb{R}^2$. Most importantly, the simplicial complex can capture topological information about the original data. We shall outline this process below.

Given a set of data X and a metric space Z, choose a reference map $f : X \to Z$, which we call a filter. Often a data analyst has a filter that he or she would like to study, but there are several examples of filters that have been useful in the past when using the mapper process. A density estimator is one such example. The most common choice of metric space is $Z = \mathbb{R}$.

Next, we choose a covering $\mathcal{U}$ of Z. If $Z = \mathbb{R}$, one common covering is $\mathcal{U}(R, e)$, which is defined as follows: Let R and e be positive real numbers. Then the covering $\mathcal{U}(R, e)$ consists of all intervals of the form $U_k := [kR - e, (k + 1)R + e]$, where k is a non-negative integer.

Suppose our cover is $\{U_\iota\}_{\iota \in I}$. Our next step is to construct the sets $X_\iota := f^{-1}(U_\iota)$. Choose $\varepsilon > 0$ and, within each X_ι, find the clusters using single-linkage clustering with distance parameter ε. Note that now X has been parametrized by the pairs of the form (ι, c), where ι is in the index set I, and C is a cluster in X_ι.

Finally, construct the formal simplicial complex with vertex set consisting of all possible pairs (ι, C), and where the vertices $(\iota_0, C_0), (\iota_1, C_1), \cdots, (\iota_n, C_n)$ span an n-simplex if and only if $\cap_{i=1}^n C_i \neq \emptyset$. This simplicial complex can usually plotted in $\mathbb{R}^2$ and can give insight to the data's structure. Figure 5, taken from [?] demonstrates the mapper process.

1.4 Probability of Random Spaces

Any statistical method is useless without a firm probabilistic background. For example, in order to use persistent homology, we must be able to answer the following question: How long must a generator of H_n persist in order to conclude with any degree of confidence that this generator represents genuine structure

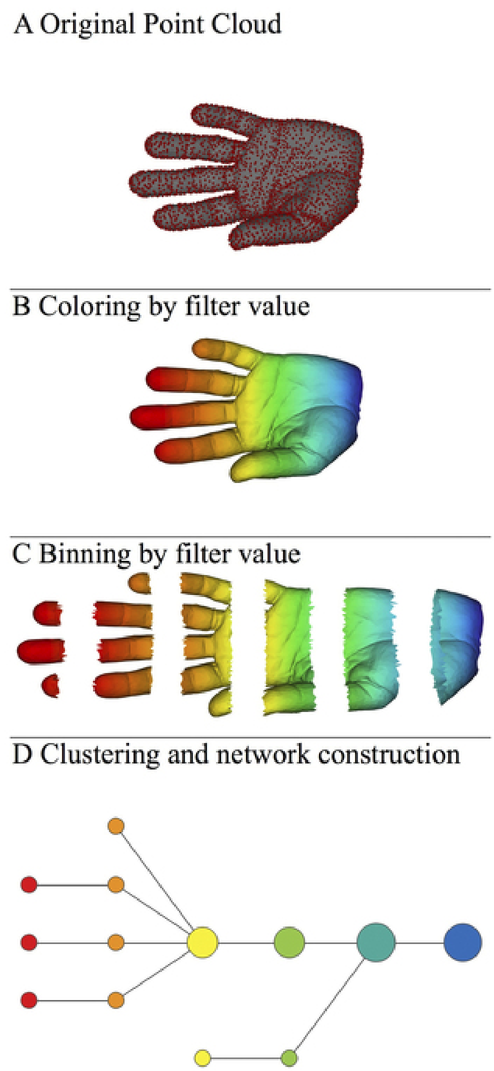


Figure 5: A demonstration of the mapper process

in the data?

There are several different approaches to solving this question. One approach attempts to develop persistent homology methods that relies on random sampling with replacement. Fatsy et. al [?], and Chezal et. al [?] have proved the validity of the bootstrap method to create confidence bands for persistence diagrams that use the kernel density estimator, distance to measure, or kernel distance.

Another approach attempts to study the space of all persistent diagrams. For example, in [?], Blumberg et. al showed that one can create confidence intervals in persistent homology by considering the space of all persistence diagrams.

Still another approach attempts to study the distribution of possible spaces generated by data under certain conditions. Several mathematicians are working on investigating the properties of such spaces. For example, Kahle has investigated the asymptotic properties of random geometric complexes in [?] and of random clique complexes in [?].

While much is known about the homological group structure of random topological spaces, not much is known about how cohomology operations behave on such random spaces. This thesis proves some results about these operations. In Chapter 2, we explore some properties of cohomology operations on certain random topological spaces. In Chapter 4, we explore the idea of a random cohomology operator in a more algebraic setting. It is assumed throughout that the reader has a good understanding of algebraic topology. There are several good textbooks on algebraic topology and cohomology operations that one can use for reference, such as [?][?].

CHAPTER 2 RANDOM COMPLEXES

2.1 Introduction

The study of the probabilistic properties of topological spaces generated by data has its roots in the study of random complexes. Two frameworks for studying random complexes are random clique complexes and random geometric complexes. Random clique complexes have the advantage that they have a simple set up, while random geometric complexes have the advantage that they better represent data in a metric space. Random geometric complexes will be discussed in more detail throughout this chapter. We shall briefly discuss random clique complexes in this section. For a more in depth presentation of the subject, one can refer to Bollobás' book [?] for details.

Let X be a set of n points and p be between 0 and 1. A random graph with parameter p is a graph with vertex set X and for any two points x and y in X, an edge between x and y with probability p. From such a random graph, one can create a random clique complex using the Vietoris-Rips complex. That is, let an n simplex span the n vertices $x_0, x_1, x_2, \ldots, x_n$ if and only if each pair of these vertices is connected by an edge.

Should one wish to study the asymptotic properties of a random graph, one may set $p = p(n)$ as a function of n in order to investigate the asymptotic topological properties of a random clique complex with parameter $p(n)$ as n goes to infinity. Kahle [?] has provided some interesting results about the asymptotic properties of the homology groups under this framework. Clique complexes are sometimes called flag complexes.

The probabilistic framework of random clique complexes is useful in areas like random networks, but it is not appropriate for studying the shape of data that lie in a metric space, because all notion of distance is lost by randomly assigning edges between points. The framework of random geometric complexes allows us to study the properties of data with some notion of distance.

Let X be a set of n points in a metric space with metric d chosen independent and identically distributed according to the density f. Let $r > 0$. The random geometric graph on X with parameter r is the graph with vertices X and an edge between any two points x_1 and x_2 of X whenever $d(x_1, x_2) < r$. A more complete exposition of geometric random graphs is given in [?].

From this random graph, we can create a random geometric complex using, for example, the Vietoris-

Rips complex. Should one wish to study the asymptotic properties of a random geometric complex, one may set $r = r(n)$ as a function of n. Then, under different conditions on r, one may investigate the topological properties of a random graph as n goes to infinity.

This framework takes into account the notion of distance between the points of X, so it is appropriate to consider when one is interested in data in a metric space. Kahle [**?**] has also proved several results about asymptotic properties of the homology groups under this framework. We shall go into the details of Kahle's results in Section 2.3.

When investigating a space's topology, one would ideally compute the space's homotopy type. However, this is often very difficult. So topologists often start by computing a space's homological or cohomological group structure. While homology groups are easier to compute, they do not contain as much information about a space's topology as a space's homotopy structure. This has led topologists to develop additional tools that give more information about a space's topology than its homological group structure alone.

Cohomology operations on a space are such one example of such tools that give additional information about a space's topological structure that may be contained in neither its homological nor cohomogical group structure. In particular, a cohomology operation is a family of functions between cohomology groups that satisfy certain properties, and, most importantly, is a topological invariant of a space. Thus, cohomology operations are useful tools for distinguishing spaces that are not homotopic. For example, with $\mathbb{Z}/2$ coefficients $S^1 \vee S^2$ and $\mathbb{RP}^2$ have the same cohomology groups, but one can find a cohomology operation – the Bockstein – that is different on these two spaces.

That one can use cohomology operations to distinguish between topological spaces would make them useful in understanding the underlying distribution of a random space or the shape of data. To apply these tools with any degree of certainty, one must understand the asymptotic properties of cohomology operations. The goal of this chapter is to provide some information on the asymptotic properties of the rank of cohomology operations on certain random spaces.

2.2 Definitions

One model for a random topological space is the geometric random graph. Such a model is useful when discussing the shape of random points in a metric space. As a stepping stone we shall first introduce geometric graphs, which have no randomness.

Definition 2.1. Let X be a subset of $\mathbb{R}^d$ and $r > 0$. Define the geometric graph $G(X; r)$ as the graph with

1. vertices X, and

2. edges between vertices x and y whenever $d(x, y) < r$.

Thus a geometric graph is dependent on X and on the parameter r. We see that if r is sufficiently large, the graph is completely connected, while if r is sufficiently small, the graph is completely disconnected. To add randomness to this construction, we choose the points of X randomly. It is important to note that while X may lie in $\mathbb{R}^d$, the geometric graph $G(X; r)$ is an abstract set of points and edges that does not lie in Euclidean space.

Definition 2.2. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a probability density function. Let $x_1, x_2, \ldots$ be a sequence of i.i.d d-dimensional random variables with density f. Let $X_n = \{x_1, x_2, \ldots, x_n\}$. For $r > 0$, the graph $G(X_n; r)$ is said to be a random geometric graph.

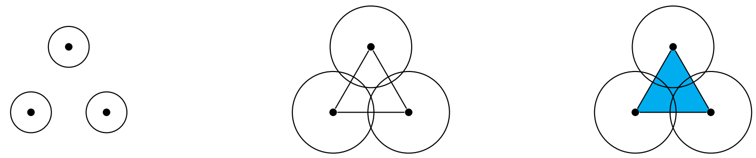
One usually regards $r = r_n$ as a function of n. We are then interested in studying the asymptotic behavior of random graphs, as well as the asymptotic behavior of the cohomology operations on the cohomolgy ring generated by random graphs, as n tends to ∞.

A random graph also lays the groundwork for a random topological space in the following way: One may view a random graph as the 1-skeleton of a random simplicial complex if, in addition to a random graph, we have a rule that determines which k-dimensional faces ($k \geq 2$) to include in the complex. The following definitions give two different ways to construct a random simplicial complex from a random graph.

Definition 2.3. The random Čech complex $\check{C}(X_n, r)$ on the set X_n with parameter r is the simplicial complex with vertex set X_n, and σ a face whenever

$$\bigcap_{x_i \in \sigma} B(x_i, r/2) \neq \emptyset.$$

One disadvantage of the Čech complex is that it is computationally expensive to compute. One must store the distance between all vertices as well as simplicies of various dimensions. In topological data analysis, one is interested in the complexes that are generated as r varies over a range of value. Creating Čech complexes can be prohibitively expensive for large data sets. Thankfully, there is another construction that we can use



$V(X;0.27)$ $\quad$ $\check{C}(X;0.54)$ $\quad$ $V(X;0.54)$

Figure 6: A demonstration of the difference between the Čech and Vietoris-Rips complexes

that is much less computationally expensive.

Definition 2.4. The random Vietoris-Rips complex $V(X_n, r)$ on the set X_n with parameter r is the simplicial complex with vertex set X_n and with σ as face whenever

$$d(x_i, x_j) < r,$$

for every pair $x_i, x_j \in \sigma$.

To create the Vietoris-Rips complex, one need only know the complex's one-skeleton. It is therefore much less computationally expensive to compute than the Čech complex, and so it is used most often in topological data analysis.

Let X be a set of vertices and r be greater than 0. Suppose a k-face σ occurs in a $\check{C}(X, r)$. Then $\cap_{x_i \in \sigma} B(x_i, r/2) \neq 0$. In particular, for any pair of vertices x_i and x_j of σ, we have $d(x_i, x_j) < r$. Thus σ is in $V(X, r)$. This proves that $\check{C}(X, r)$ is a subset of $V(X, r)$. In fact, we have the following preposition that relates the Čech and Vietoris-Rips complexes. One can find a proof of this proposition in [**?**]. Figure 6 gives an example to show that the $\check{C}(X, r)$ need not equal $V(X, r)$.

Proposition 2.5. Let X_n be a set of n points in a metric space and let $r > 0$. Then

$$V(X_n, r) \subseteq \check{C}(X_n, 2r) \subseteq V(X_n, 2r).$$

For example, let $X = \left\{(0,0), \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), (1,0)\right\}$ as a subset of $\mathbb{R}^2$. If $r = 0.54$, then $V(X, r) = X$, which is a zero-complex, $\check{C}(X, r)$ is a one-complex, and $V(X, r)$ is a two-complex. We illustrate this example in Figure 6.

2.3 Random Geometric Complexes

Random geometric complexes are a probabilistic model for investigating the shape of spaces generated by data. Our goal is to investigate the asymptotic properties of the topology of random geometric complexes. To do so, we need to be familiar with the language of asymptotics.

Definition 2.6. We have the following definitions for asymptotic behavior.

- $g(n) = O(h(n))$ means that there exists n_0 and k such that for $n > n_0$, we have $g(n) \leq k \cdot h(n)$.
- $g(n) = \Omega(h(n))$ means that there exists n_0 and k such that for $n > n_0$, we have that $g(n) > k \cdot h(n)$.
- $g(n) = \Theta(h(n))$ means that $g(n) = O(n)$ and $g(n) = \Omega(n)$.
- $g(n) = o(h(n))$ means that for every $\varepsilon > 0$, there exists n_0 such that for $n > n_0$, we have $g(n) \leq \varepsilon \cdot h(n)$.
- $g(n) = \omega(h(n))$ means that for every $k > 0$, there exists n_0 such that for $n > n_0$, we have that $g(n) \geq k \cdot h(n)$.

We should like to be able to say something about the asymptotic properties of the cohomology operations on random spaces. It turns out that we can say more about the asymptotics of feasible graphs than of arbitrary graphs.

Definition 2.7. A connected graph is $\mathbb{R}^d$-feasible if it is geometrically realizable as a geometric graph in $\mathbb{R}^d$.

Consider the graph in $\mathbb{R}^2$ consisting of eight vertices labeled $x_1, x_2, \ldots, x_8$ and with edge set consisting exactly of edges connecting vertex x_8 to all other vertices. This is known as the complete bipartite graph $K_{1,7}$. Note that it must be the case that the distance between some pair of the first seven vertices must be less than the maximum of the distances between the eighth vertex and all other vertices. That is

$$\min_{1 \leq i < j \leq 7} d(x_i, x_j) \leq \max_{1 \leq i \leq 7} d(x_i, x_8).$$

This implies that $K_{1,7}$ is not feasible. Figure 7 gives a picture of $K_{1,7}$.

A theorem of Penrose will help us prove results about the asymptotics of random spaces. To use this theorem, we must understand the difference between an arbitrary subgraph and an induced subgraph.

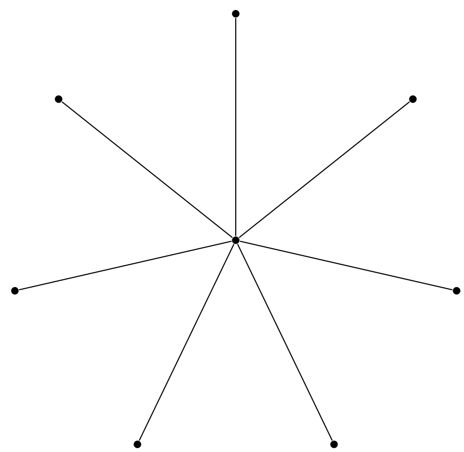


Figure 7: The graph $K_{1,7}$, which is not $\mathbb{R}^2$-feasible

Definition 2.8. Let G be a graph. The graph H is a subgraph of G whenever the vertices and edges of H are also in G. The graph K is an induced subgraph of G whenever the set of vertices of K are vertices of G and an edge exists in K only if it exists in G.

For example, consider the complete graph G on the vertex set $\{a, b, c, d\}$. Then the graph with vertex set $\{a, b, c\}$ and edges set $\{(a, b), (b, c)\}$ is a subgraph of G, but not an induced subgraph, because any induced subgraph of G with the vertices a and c must contain the edge (a, c).

Let H be a graph. Let $G_n(H)$ denote the number of induced subgraphs of $G(X_n; r)$ that are isomorphic to H. Let $J_n(H)$ denote the number of components of $G(X_n; r)$ that are isomorphic to H.

For a feasible subgraph H of order k, and $\mathcal{Y} \in (\mathbb{R}^d)^k$, define the indicator function on $I_H(\mathcal{Y})$ on sets $\mathcal{Y}$ of k elements in $\mathbb{R}^d$ by $I_H(\mathcal{Y}) = 1$ whenever the geometric graph $G(\mathcal{Y}, 1)$ is isomorphic to H, and 0 otherwise. Let

$$\mu_H = k!^{-1} \int_{\mathbb{R}^d} f(x)^k dx \int_{(\mathbb{R}^d)^{k-1}} I_H(\{0, x_1, x_2, \ldots, x_{k-1}\}) d(x_1, x_2, \ldots, x_{k-1}).$$

Penrose proved the following in [**?**].

Theorem 2.9. (Penrose)

Let $\lim_{n\to\infty} r = 0$ and H be a connected feasible graph of order $k \geq 2$. Then

$$\lim_{n\to\infty} \frac{E(G_n(H))}{r^{-d(k-1)}n^{-k}} = \lim_{n\to\infty} \frac{E(J_n(H))}{r^{-d(k-1)}n^{-k}} = \mu_H$$

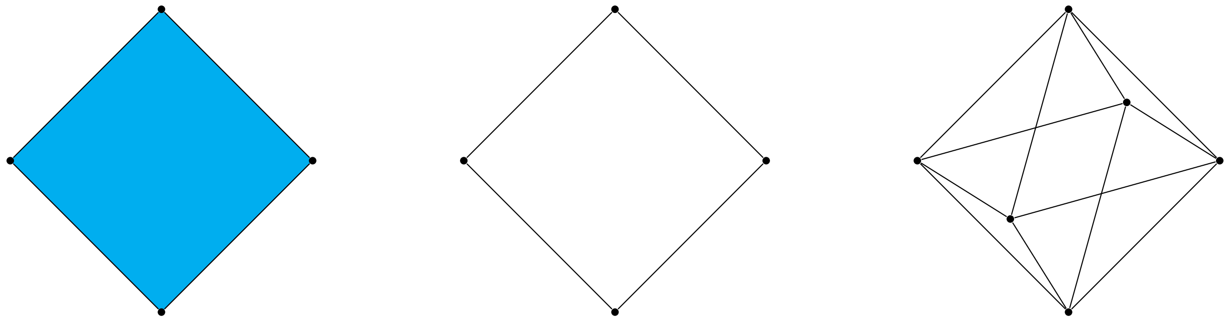


Figure 8: The 2-dimensional cross-ploytope, its boundary O_1, and the 1-skeleton of O_2.

Many of the theorems in this chapter are derived from Penrose's result. We should like to draw particular attention to the fact that this theorem is only known to be true when H is a connected feasible graph. This will not restrict Kahle's results on the asymptotics of Betti numbers, but it will restrict our results when we investigate the asymptotics of cohomology operations.

Definition 2.10. For a topological space X, in integral homology, the k-th Betti number β_k is the number of $\mathbb{Z}$ summands in the k-th homology group of X.

We shall see that the expectation of the k-th Betti number depends asymptotically on the number of induced subgraphs that are isomorphic to the 1-skeleton of the smallest simplicial complex that supports k-th homology. This smallest complex is what we define as O_k in the definition below.

Definition 2.11. Let $e_1, e_2, \ldots, e_{k+1}$ be the standard basis vectors in $\mathbb{R}^{k+1}$. The $(k+1)$-dimensional cross-polytope is the convex hull of the $2k+2$ points $\{\pm e_i\}$. Its boundary is a k-dimensional simplicial complex, which we denote by O_k.

Figure 8 gives an illustration of the 2-dimensional cross polytope, its boundary O_1, as well as the 1-skeleton of O_2.

To prove results about the asympototics of the Betti numbers of a random complex, we must count the subcomplexes that support non-trivial homology. The following lemma will be useful in this regard and is proved in [**?**].

Lemma 2.12. If Δ is a flag complex, then any non-trivial element of $H_k(\Delta)$ is supported on a subcomplex C of Δ with at least $2k+2$ vertices. In addition, if C has exactly $2k+2$ vertices, then C is isomorphic to O_k.

The asymptotics of the Betti numbers of random geometric complexes depend on the rate at which the parameter r_n goes to 0. There are three regimes of interest: The subcritical regime, the critical regime, and the supercritical regime. We provide the results for the subcritical and critical regime in Theorems 2.13 and 2.14. We shall employ the proof methods of these results to cohomology operations in Section 2.4.

The first case we shall cover is the subcritical regime. It is interesting to note that these results hold even if $d > k$. We duplicate the proof for the case $n \geq 2$, but we do not provide much exposition. We refer the reader who is interested in additional details of this proof, as well as results for the asymptotics of the Betti numbers in the supercritical regime, to Kahle's paper [?] for further details.

Theorem 2.13. (Kahle)(Subcritical Regime) Let $d \geq 2$, $k \geq 1$, $\varepsilon > 0$, and $r_n = O(n^{-1/d-\varepsilon})$. Let f be a bounded and measurable density function on $\mathbb{R}^d$. Let X_n be a set of n points chosen independent and identically distributed from f. Then

$$\frac{E[\beta_k]}{n^{2k+2}r^{d(2k+1)}} \to C_k,$$

as $n \to \infty$, where C_k is a constant that depends only on k and the underlying density function f.

Proof. For an arbitrary simplicial complex Δ, let $o_k(\Delta)$ be the number of induced subgraphs of Δ that are combinatorially isomorphic to the 1-skeleton of O_k. Let $\tilde{o}_k(\Delta)$ be the number of components of Δ that are combinatorially isomorphic to the 1-skeleton of the cross-polytope O_k. Finally, let $f_k^{\geq i}(\Delta)$ denote the number of k-dimensional faces on connected components containing at least i vertices.

We have the following inequality

$$\tilde{o}_k \leq \beta_k \leq \tilde{o}_k + f_k^{\geq 2k+3}, \tag{1}$$

which follows immediately from Lemma 2.12. We should like to further overestimate f_k^{2k+3}. For each k-dimensional faces in a component with at least $2k+3$ vertices, extend to a connected subgraph with exactly $2k+3$ vertices and $\binom{k+1}{2}+k+2$ edges. There are c_k ways to do this, where c_k is a constant that depends only on k. For $i \in \{1, 2, \ldots, c_k\}$, let s_i count the number of subgraphs isomorphic to graph i for some indexing of the c_k graphs. Then we have $f_k^{2k+3} \leq \sum_{i=1}^{c_k} s_i$. Taking expectations of the inequalities in 1 yields

$$\begin{aligned} E[\tilde{o}_k] \leq E[\beta_k] &\leq E[\tilde{o}_k] + E[f_k^{\geq 2k+3}] \\ &\leq E[\tilde{o}_k] + E\left[\sum_{i=1}^{c_k} s_i\right] \\ &\leq E[\tilde{o}_k] + \sum_{i=1}^{c_k} E[s_i]. \end{aligned}$$

We have $E[s_i] = O\left(n^{2k+3} r^{(2k+2)d}\right)$ and $E[\tilde{o}_k] = \Theta\left(n^{2k+2} r^{(2k+1)d}\right)$. By assumption, $nr^d \to 0$ as $n \to \infty$. Thus, we have that

$$\frac{E\left[\beta_k\right]}{E\left[\tilde{o}_k\right]} \to 1,$$

and thus

$$E\left[\beta_k\right] = \Theta\left(n^{2k+2} r^{(2k+1)d}\right).$$

□

We note that a similar theorem with a similar proof exists for the Čech complex. We again refer the reader to [**?**] for details. The theorem for the critical regime follows very quickly from Theorems 2.9 and 2.13.

Theorem 2.14. (Kahle)(Critical Regime) For either the random Vietoris-Rips complex or Čech complex on a probability distribution on $\mathbb{R}^d$ with bounded measurable density function f, if $r = \Theta\left(n^{-1/d}\right)$ and $k \geq 1$ is fixed, then $E[\beta_k] = \Theta(n)$.

Proof. By the same reasoning in the proof of Theorem 2.13, we have

$$E[\tilde{o}_k] \leq E[\beta_k] \leq E[\tilde{o}_k] + E[f_k^{\geq 2k+3}].$$

Theorem 2.9 yields $E[\tilde{o}_k] = \Theta(n)$ and $E[f_k^{\geq 2k+3}] = O(n)$, and the desired result follows. □

2.4 Subcritical and Critical Regimes

In addition to investigating a space's homological or cohomological group structure, one can examine cohomology operations on a space's cohomology groups to better understand its topology. We provide the definition of a cohomology operation here, and refer the reader to any number of textbooks or papers on the subject, such as [**?**], for a more in-depth exploration.

Definition 2.15. Let G_1 and G_2 be groups, and p and q be nonnegative integers. A cohomology operation of type $(G_1, p; G_2, q)$ is a family of functions $\theta_Y : H^p(Y; G_1) \to H^q(Y; G_2)$, one for each space Y, satisfying the naturality condition $f^\star \theta_Z = \theta_Y f^\star$.

We remark that our definition of cohomology operations depends on p and q and so the definition given here is for unstable operations. In Section 2.3 we saw that the expectation of the k-th Betti number for a simplicial complex depends asymptotically on $\tilde{o}_k$ – the number of components of Δ combinatorially isomorphic to the 1-skeleton of the cross-polytope O_k. We should like an analogous construction for cohomology operations.

Definition 2.16. For a cohomology operation θ, define a space Y to be θ-essential if $\theta : H^p(Y; G_1) \to H^q(Y; G_2)$ is non-trivial. Define a simplicial complex X to be θ-minimal if it is θ-essential and all complexes with fewer vertices have trivial θ. Define $m(\theta)$ as the number of vertices in a θ-minimal complex.

Certainly $m(\theta)$ will exist for any cohomology operation, but there is no reason to believe that the complex with fewest vertices that supports θ is unique. The following lemma gives an example for computing $m(\theta)$.

Lemma 2.17. The smallest flag complex in that supports the Bockstein

$$\beta : H^1(X; \mathbb{Z}/2) \to H^2(X; \mathbb{Z}/2)$$

has at most thirteen vertices. That is, $m(\beta) \leq 13$.

Proof. Figure 2.4 gives us visualization of $\mathbb{RP}^2$ in $\mathbb{R}^2$. The vertices with the same index on opposite sides of the octagon are identified, and the edges between these vertices are identified as indicated by the arrows.

□

Lemma 2.17 gives the perhaps the simplest example of computing $m(\theta)$ for some cohomology operation θ. For an arbitrary cohomology operation θ, the number $m(\theta)$ is unknown. Ideally, we should like to use $m(\theta)$ to find a bound for the rank of a cohomology operation, but in order to apply Theorem 2.9, we must work with feasible graphs. Thus, we require the following definition.

Definition 2.18. Let $\mathcal{F}_{\theta,d}$ be the collection of all $\mathbb{R}^d$-feasible θ-essential spaces. Let $m'(\theta, d)$ be the minimum number vertices of a complex in $\mathcal{F}_{\theta,d}$.

Finding feasible spaces that support a given cohomology operation seems to be a very difficult problem. We have not shown that $\mathcal{F}_{\theta,d}$ is non-empty, but our intuition suggests that this is not the case. If $\mathcal{F}_{\theta,d}$

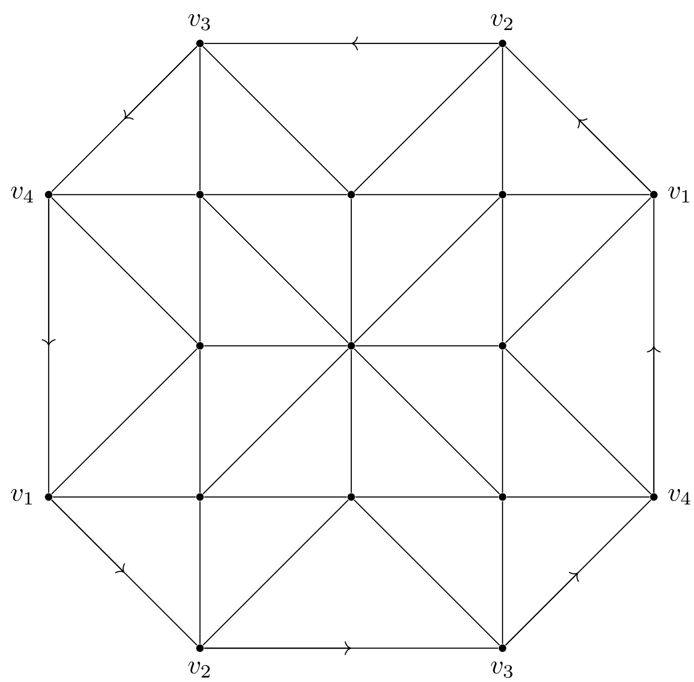


Figure 9: $\mathbb{RP}^2$ as a flag complex

were empty, then $\theta : H^p(G(X_n; r_n)) \to H^q(G(X_n; r_n))$ would be trivial for all n. In this case, studying the asymptotic properties of θ for this particular d is not interesting.

We are now ready to give our first result on the asymptotic properties of a cohomology operation.

Theorem 2.19. (Subcritical Regime) Assume that $\mathcal{F}_{\theta,d}$ is non-empty. Let θ be a cohomology operation. Let f be a bounded and measurable density function on $\mathbb{R}^d$. Let X_n be n points chosen independent and identically distributed from f. For $d \geq 2$, $k \geq 1$, $\varepsilon > 0$, and $r_n = O\left(n^{-1/d-\varepsilon}\right)$, the expectation of the rank of a cohomology operation θ of the random Vietoris-Rips complex $V(X_n; r_n)$ satisfies

$$\frac{E\left[\operatorname{rank}(\theta : H^p(G(X_n; r_n)) \to H^{p+k}(G(X_n; r_n))\right]}{n^{m'(\theta,d)} r^{d(m'(\theta,d)-1)}} \to C_k,$$

where C_k is a constant that depends on f, k and θ.

Proof. Let $\tilde{q}_k$ be the number of feasible components of Δ that are isomorphic to a θ-minimal complex. Let $f_k^{\geq i}$ be the number of k-dimensional faces on connected components of Δ that contain at least i vertices. Then we have

$$\tilde{q}_k \leq \operatorname{rank}(\theta, d) \leq c_k \tilde{q}_k + f_k^{\geq m'(\theta,d)+1},$$

where c_k is a constant. We can overestimate $f_k^{\geq m'(\theta)+1}$ as in the proof of Theorem 2.13. Then, taking expectations of both sides, we have

$$E[\tilde{q}_k] \leq E[\text{rank}(\theta)] \leq c_k E[\tilde{q}_k] + O\left(n^{m'(\theta,d)+1} r^{m'(\theta,d)d}\right).$$

By Theorem 2.9, we have

$$E[\tilde{q}_k] = \Theta\left(n^{m'(\theta,d)} r^{(m'(\theta,d)-1)d}\right).$$

Since $nr^d \to 0$, we have that $E[\text{rank}(\theta)]/E[\tilde{q}_k] \to 1$, and the result follows. □

We could also apply the same techniques to prove an analogous result about the rank of cohomology operations on a Čech complex.

Theorem 2.20. (Critical Regime) Assume that $\mathcal{F}_{\theta,d}$ is non-empty. Let θ be a cohomolgy operation on a random Vietoris-Rips or Čech complex with probability distribution on $\mathbb{R}^d$ whose density function is bounded. If $r = \Theta\left(n^{-1/d}\right)$, then $\text{rank}(\theta) = \Theta(n)$.

Proof. Since Penrose's result extends to the case when $r = \Theta\left(n^{-1/d}\right)$, the proof is the same as in the case of the subcritical regime. □

CHAPTER 3 MINIMAL COMPLEXES

3.1 Introduction

In Section 2.3, we reviewed some results on the asymptotic properties of the expectation of Betti numbers on random spaces. We have seen that, asymptotically, the expectation of the k-th Betti number for a random geometric complex is dependent on the number of components whose 1-skeleton is combinatorially isomorphic to O_k – the boundary of the $(k+1)$-st dimensional cross-polytope.

The complex O_k is a well-studied object. In particular, we know how many vertices are in O_k, that O_k is the smallest complex that supports non-trivial k-homology, and that O_k is geometrically feasible in $\mathbb{R}^2$. These facts allow one to apply Theorem 2.9 and to find expectations in Theorems 2.13 and 2.14.

In Section 2.4, we derived analogous results about the asymptotic properties of the expected value of a cohomology operation's rank on a random geometric complex. Let θ be a cohomology operation and d be a positive integer. We have seen that, asymptotically, the expectation of the rank of θ relies on the number of feasible components that are isomorphic to an $\mathbb{R}^d$-feasible θ-minimal complex. Recall that $m'(\theta, d)$ denotes the number of vertices in such a component and was an integral part of the result in Theorem 2.19.

However, $\mathbb{R}^d$-feasible θ-minimal complexes have not been well-studied, and finding $m'(\theta, d)$ exactly appears to be a very difficult problem. In this chapter, we find upper bounds on $m'(\theta, d)$ for certain squaring operations.

3.2 The complex O_k

Recall from Definition 2.11 that O_k is the boundary of the $(k+1)$-dimensional cross polytope, which is defined as the set $\{x \in \mathbb{R}^{k+1} : \|x\|_1 \leq 1\}$. These O_k are important as, for each k, it is the complex with fewest vertices and nontrivial k-th homology. See [**?**], for example, for a proof of this fact. In this section, we shall count the number simplices in O_k.

Lemma 3.1. The total number of simplices of all degrees in O_k is $3^{k+1} - 1$.

Proof. Let e_i be the i-th standard basis vector in $\mathbb{R}^{k+1}$. Then the $2k+2$ vertices of O_k are $\pm e_i$. A simplex of O_k can be thought of as a list with $k+1$ ordered entries, where the i-th entry is either e_i, $-e_i$, or left empty. The only possibility that is not allowed is if all entries are left empty, because the empty set is not a simplex. Thus, we see that the total number of simplicies in O_k is $3^{k+1} - 1$. □

3.3 Subdivision of Δ-complexes

We are primarily interested in flag complexes, but it will be useful to start with Δ-complexes. We shall see that a Δ-complex, when sufficiently subdivided, is a flag complex. The following defines a Δ-complex and a regular Δ-complex.

Definition 3.2. Let

$$\Delta^n = \{(t_0, \ldots, t_n) \in \mathbb{R}^{n+1} | \sum_i t_i = 1 \text{ and } t_i \geq 0 \text{ for all } i\}.$$

A Δ-complex structure on a space X is a collection of maps $\sigma_\iota : \Delta^n \to X$, with n depending on the index ι, such that

(i) The restriction to the interior $\sigma_\iota|\mathring{\Delta}^n$ is injective, and each point of X is in exactly one such restriction.

(ii) Each restriction of σ_ι to a face of Δ^n is one of the maps $\sigma_\alpha : \Delta^{n-1} \to X$, and this restriction preserves the ordering of the vertices.

(iii) A set $A \subset X$ is open if and only if $\sigma_\iota^{-1}(A)$ is open in Δ^n for each σ_ι.

A Δ-complex in which every map σ_ι is injective is called a regular Δ-complex.

An n-simplex σ can be subdivided into $(n+1)!$ n-simplices via barycentric subdivision. Details of this subdivision can be found in standard algebraic topology textbooks [**?**]. In particular, if τ is one of the simplices in the barycentric subdivision of the n-simplex σ, then τ is an injection, the image of τ contains at most one zero-simplex of Δ^n, and the following composition holds:

$$\sigma\tau : \Delta^n \hookrightarrow \Delta^n \to X$$

Lemma 3.3. The subdivision of any Δ-complex is a regular Δ-complex.

Proof. Let σ be a n-simplex of a Δ-complex on X. Let τ be a simplex in the barycentric subdivision of σ. Then $\sigma\tau : \Delta^n \hookrightarrow \Delta^n \to X$. Suppose that $\sigma\tau(x) = \sigma\tau(y)$. We must show that $x = y$.

Note that both $\tau(x)$ and $\tau(y)$ must each lie in the interior of some, possibly non-proper, subface of Δ^n. So, for some non-negative integers m and k that are less than n, we know that $\tau(x)$ lies in the interior of Δ^m and $\tau(y)$ lies in the interior of Δ^k. Further, we know that both Δ^m and Δ^k are subfaces of Δ^n.

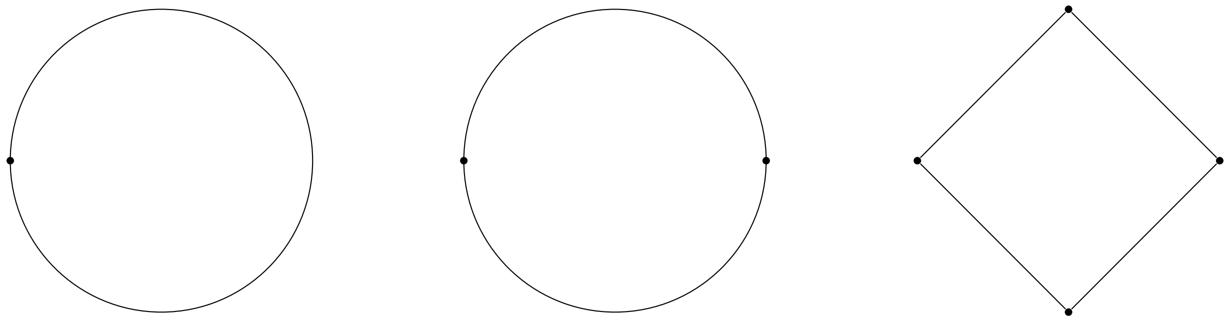


Figure 10: Subdivisions of a Δ-complex

Let σ' and σ'' be the restrictions of σ to Δ^m and Δ^k, respectively. Then

$$\sigma'\tau(x) = \sigma\tau(x) = \sigma\tau(y) = \sigma''\tau(y).$$

By Condition (i) of Definition 3.2, we know that $m = k$. Thus $\sigma' = \sigma''$ as functions on Δ^m. Further, since τ is a simplex in the barycentric subdivision of σ we know that $\tau(x)$ and $\tau(y)$ lie in the interior of the same m-subface of Δ^n. Therefore, $\tau(x)$ and $\tau(y)$ both lie in the interior of Δ^m. Since $\sigma'\tau(x) = \sigma'\tau(y)$, by Condition (i) of Definition 3.2, we know that $\tau(x) = \tau(y)$, and since τ is injective, $x = y$. □

The following corollary follows immediately from Lemma 3.3.

Corollary 3.4. If a space X has a model as a Δ-complex, then it has a model as a regular Δ-complex.

We wish to work with flag complexes. The following lemma, which we supply without proof, shows how to construct a flag complex from a regular Δ-complex.

Lemma 3.5. The subdivision of any regular Δ-complex is a flag complex.

Note that regularity is necessary in Lemma 3.5. For example, the space with one 1-simplex whose boundary points are identified is a Δ-complex, but not a regular Δ-complex. Subdividing this complex once gives a regular Δ-complex that is not a flag complex. Subdividing the regular Δ-complex yields a flag complex. Figure 3.3 gives a visualization of these subdivisions.

Lemma 3.5 and Lemma 3.3 together give the following corollary.

Corollary 3.6. If a space X has a model as a Δ-complex, then it has a model as a flag complex.

3.4 An Upper Bound for $m'(Sq^{2^i}, d)$

The Steenrod squares, $Sq^j : H^\ell(X;\mathbb{Z}/w) \to H^{\ell+j}(X;\mathbb{Z}/2)$ are important examples of stable cohomology operations. One important property of the Steendrod squares is that the operations Sq^{2^i} are indecomposable, while all other Steenrod squares are decomposable. So, to understand the Steenrod squares, it is important to understand the Sq^{2^i}. In this section, we shall prove an upper bound for $m'(Sq^{2^i}, d)$.

We first note that $Sq^{2^i}(x) = 0$ whenever the degree of x is less than 2^i. Thus, if $\ell < 2^i$, then $Sq^{2^i} : H^\ell(X) \to H^{\ell+2^i}(X)$ will be trivial, no matter how many vertices are in X. In this case,

$$m'\left(Sq^{2^i} : H^\ell \to H^{\ell+2^i}, d\right) = \infty.$$

We are interested in finding bounds $m'(Sq^{2^i}, d)$ in the non-trivial case. It will suffice to consider

$$Sq^{2^i} : H^{2^i+n} \to H^{2^{i+1}+n}$$

for $n \geq 0$.

Theorem 3.7. We have the following bound

$$m'(Sq^{2^i} : H^{2^i+n} \to H^{2^{i+1}+n}, d) \leq 2n + k_{2^i,d},$$

where $k_{2^i,d} := m'(Sq^{2^i} : H^{2^i} \to H^{2^{i+1}}, d)$.

Proof. Let X be a complex on k_{2^i} vertices that is Sq^{2^i}-essential. The two point suspension of X is the complex ΣX created by adding two vertices of x and y to X with edges between x and every vertex of X and edges between y and every vertex of X, but no edge between x and y. Then ΣX is Sq^{2^i}-essential, because Sq^{2^i} is a stable operation. The assertion now follows by induction on n. $\square$

We note here that it seems possible that the bound in Theorem 3.7 is sharp.

Next, let $k_{2^i,d} = m'(Sq^{2^i} : H^{2^i} \to H^{2^{i+1}})$. We shall find bounds on $k_{2^i,d}$. Certainly $\mathbb{RP}^{2^{i+1}}$ is Sq^{2^i}-essential. Thus, to find an upper bound on k_{2^i}, it suffices to construct a flag complex that is homotopic to $\mathbb{RP}^{2^{i+1}}$ and count its vertices. In the following theorem, it is necessary that d be large enough so that our constructed flag complex of $\mathbb{RP}^{2^{i+1}}$ is $\mathbb{R}^d$-realizable.

Lemma 3.8. For a sufficiently large d, there is a flag complex representation of $\mathbb{RP}^{2^{i+1}}$ with $\frac{3^{2^{i+1}}-1}{2}$ vertices. In particular, $k_{2^i,d} \leq \frac{3^{2^{i+1}-1}}{2}$.

Proof. We shall construct a flag complex representation of $\mathbb{RP}^k$. Begin with O_k. By identifying opposite simplices in the appropriate way, we create a regular Δ-complex representation of $\mathbb{RP}^k$. If we subdivide this complex, by Lemma 3.5, we shall have a flag complex representation of $\mathbb{RP}^k$.

In order to count the number of vertices in this flag complex representation of $\mathbb{RP}^k$, we note first identifying opposite simplices of O_k and then subdividing yields the exact same space as first subdividing O_k and then identifying opposite simplices. It will be easier to count the number of vertices in this flag complex if we subdivide O_k first.

By Lemma 3.1, the number of vertices in the subdivision of O_k is $3^{k+1}-1$. Identifying opposite vertices gives us a flag complex representation of $\mathbb{RP}^k$ with $\frac{3^{k+1}-1}{2}$ vertices. The Lemma follows by taking $k = 2^{i+1}$. □

3.5 Lower bounds on the number of vertices in $\mathbb{RP}^{2^{i+1}}$

We should also like to find a lower bound on the number of vertices in a complex that is homotopic to $\mathbb{RP}^{2^{i+1}}$. Note here that we are no longer considering flag complexes. Our intuition suggests that this could also be a lower bound on $k_{2^i,d}$. The methods we employ are the same as the methods used by Arnoux and Marin [**?**] to find a lower bound on the number of vertices in a weak cohomology $\mathbb{CP}^n$.

Definition 3.9. A weak cohomology $\mathbb{RP}^n$ is a pair (X, h) consisting of:

(1) a finite simplicial complex X.

(2) a cohomology class h in $H^1(X; \mathbb{Z}/2\mathbb{Z})$ with $h^n \neq 0$.

A full subcomplex is the higher-dimensional analog of an induced subgraph, which we defined in Definition 2.8. The full subcomplexes of a weak cohomology $\mathbb{RP}^n$ will tell us how to find construct weak cohomology $\mathbb{RP}^{n-1}$ subcomplexes.

Definition 3.10. A subcomplex Δ' of Δ is full whenever any simplex of Δ, all of whose vertices are in Δ', is included in Δ'.

For a complex Δ and a vertex v of Δ, the star of v, denoted $\mathrm{st}_\Delta(v)$, is the set of simplices of Δ that have a v as a vertex. Note that $\mathrm{st}_\Delta(v)$ is not necessarily a simplicial complex. However, the smallest subcomplex of Δ that contains $\mathrm{st}_\Delta(v)$ is a full subcomplex of Δ. The following lemma will help us find a lower bound for the number of vertices in a weak cohomology $\mathbb{RP}^n$.

Lemma 3.11. Let (X, h) be a weak cohomology $\mathbb{RP}^n$. Let Y be a full subcomplex of X on which h restricts to the zero class. Let h_Y be the restriction of h to $X - Y$. Then $(X - Y, h_Y)$ is a weak cohomology $\mathbb{RP}^{n-1}$.

Proof. Recall that if two cohomology classes a and b of a space S restrict, respectively, to zero on the open sets A and B, which together cover S, then the cup product $a \cup b$ is zero.

Note that $X - Y$ and Y can be thickened to the open sets A and B, respectively, so that A is homotopic to $X - Y$, Y is homotopic to B, and so that A and B form an open cover of X. Note that $h^{n-1} \cup h = h^n$, which is not zero, because X is a weak cohomology $\mathbb{RP}^n$. Since h restricts to the zero class on Y by assumption, we know by the first sentence of this proof that the restriction h_Y^{n-1} cannot be zero on $X - Y$. $\square$

Theorem 3.12. A weak cohomology $\mathbb{RP}^n$ has at least $\frac{(n+1)(n+2)}{2}$ vertices.

Proof. Let (X, h) be a weak cohomology $\mathbb{RP}^n$ with N vertices. Since h^n is in $H^n(X)$, we know that $H^n(X)$ must be non-trivial. So there must be an n-simplex in X. Let Δ be a p-simplex of X with $p \geq n$. Then it is a contractable full subcomplex of X, and the previous lemma asserts that $(X - \Delta, h_\Delta)$ must be a weak cohomology $\mathbb{RP}^{n-1}$. The theorem then follows by induction on N, because $X - \Delta$ has $N - (p+1) \leq N - (n+1)$ vertices and

$$n + 1 + (n-1) + 1 + (n-2) + 1 + \cdots + 2 + 1 = \frac{(n+1)(n+2)}{2}.$$

$\square$

CHAPTER 4 A RANDOM BOCKSTEIN OPERATOR

4.1 Introduction

Using the tools of algebraic topology to better understand a data set is a relatively new idea with many applications. In Chapter 1, we reviewed the generalization of cluster analysis to persistent homology. This technique provides more information on the shape of a data set than traditional cluster analysis by considering the Betti numbers of topological spaces generated by data.

A natural question that arises is whether one can gather any additional information from a data set by looking at operations on the cohomology of the topological space generated by that data set. We examined some properties of cohomology operations on random spaces in Chapter 2. In this section we shall study an algebraic version of a random cohomology operation.

In general, the Bockstein homomorphism is a connecting homomorphism of cohomology groups defined on a chain complex. Ideally, we should consider the case of a chain complex of a randomly generated topological space. Unfortunately, this problem is very difficult. The length of the chain complex, each Abelian group in the complex, and each boundary map would all add complexity to this model. We shall therefore examine in this chapter a simpler algebraic version of the above problem whose only degrees of freedom are determined by a single boundary map.

Let V and W be free-modules with coefficients in $\mathbb{Z}/p^2$. We then have the following short exact sequences

$$0 \to pV \hookrightarrow V \twoheadrightarrow \overline{V} \to 0 \quad \text{and} \quad 0 \to pW \hookrightarrow W \twoheadrightarrow \overline{W} \to 0,$$

where $\overline{V}$ and $\overline{W}$ are the reductions of V and W modulo p. Given a map $\phi : V \to W$, define ψ from $\overline{V}$ to $\overline{W}$ to be the map induced by ϕ. The Bockstein homomorphism induced by ϕ is then a map from $\ker\psi$ to $\operatorname{coker}\psi$. We give the construction of the Bockstein homomorphism for this case in more detail in Section 4.4.

Since Bockstein homomorphisms are elements of $\hom(\ker\psi, \operatorname{coker}\psi)$, it makes sense only to compare Bocksteins induced by functions from V to W that are equal modulo p. A choice of random function from V to W is the same as choosing a random m by n matrix. To this end, let ϕ be a matrix whose entries are randomly taken independent and identically distributed from the discrete uniform distribution on $\{0, 1, 2, \ldots, p^2 - 1\}$. Let ψ be the reduction of ϕ modulo p. Let β_ϕ be Bockstein homomorphism induced by ϕ. Let γ be in $\hom(\ker\psi, \operatorname{coker}\psi)$. We shall show that

$$\mathbb{P}\big(\beta_\phi = \gamma \big| \overline{\phi} = \psi\big) = \frac{1}{p^{k(n-m+k)}}.$$

In other words, we shall show that, given $\overline{\phi} = \psi$, the Bockstein homomorphisms are distributed uniformly.

4.2 Linear Algebra over $\mathbb{Z}/p^2$

Many of our calculations will be done over $\mathbb{Z}/p^2$-modules. This section reviews the theory of $\mathbb{Z}/p^2$-modules over $\mathbb{Z}/p^2$. Some of the techniques used in this section work for modules over rings other than $\mathbb{Z}/p^2$, but we shall not explore these ideas here.

Let R be a ring. Given an R-module M, we say that a subset E of M is a basis for M whenever E generates M and E is linearly independent. This definition is equivalent to the condition that every x in M can be written as a unique linear combination of elements of E with scalars in R. A module that has a basis is called a free module.

Let p be prime, and let V and W be free $\mathbb{Z}/p^2$-modules. Define

$$\overline{V} := V \bigotimes_{\mathbb{Z}/p^2} \mathbb{Z}/p \quad \text{and} \quad \overline{W} := W \bigotimes_{\mathbb{Z}/p^2} \mathbb{Z}/p.$$

So $\overline{V} = V/pV$ and $\overline{W} = W/pW$ are the reductions of V and W mod p. Note that these are $\mathbb{Z}/p$ vector spaces. For an element $x \in V$, we use $\overline{x}$ to denote its reduction modulo p. For an element y in $\overline{V}$, we use $\tilde{y}$ to denote a choice of representative in V of y, so that $\overline{\tilde{y}} = y$. Given a $\mathbb{Z}/p^2$-linear map $\phi : V \to W$, let $\overline{\phi}$ denote the induced function from $\overline{V}$ to $\overline{W}$.

Lemma 4.1. Let V be a free $\mathbb{Z}/p^2$-module. Let $p : V \to V$ be multiplication by p. Then the kernel of p is equal to the image of p.

Proof. Let $\{e_i\}$ be a basis for V. Let x be in $\ker p$. Since $\{e_i\}$ is a basis, there are α_i in $\mathbb{Z}/p^2$ such that $x = \sum_i \alpha_i e_i$. Since x is in $\ker p$ we have $px = \sum_i p\alpha_i \cdot e_i = 0$. By the independence of the e_i, we have $p\alpha_i = 0$ for each i. Thus $\alpha_i = p\beta_i$ for some $\beta_i \in \mathbb{Z}/p^2$. Then $p(\sum_i \beta_i e_i) = \sum_i \alpha_i e_i = x$. So that x is in the image of p.

Next, assume that y is in the image of p. Then there exists a $z \in V$ with $pz = y$. So $py = p^2 z = 0$. So y

is in the kernel of p. $\square$

We know that pV and $\overline{V}$ are isomorphic as $\mathbb{Z}/p$-vector spaces, because they both have the same dimension. The following lemma gives an explicit isomorphism between these two spaces.

Lemma 4.2. The map $f : pV \to \overline{V}$ defined by $px \mapsto \overline{x}$ is a $\mathbb{Z}/p$-linear isomorphism.

Proof. We show that both f and its inverse mapping g, which maps $\overline{x}$ in $\overline{V}$ to px in pV, are well-defined. To show that f is well-defined, assume that $px = py$ for some x and y in V. Then $px - py = p(x - y) = 0$. So $x - y = pz$ for some $z \in V$ by Lemma 4.1. Note that

$$\overline{x} - \overline{y} = \overline{x - y} = p\overline{z} = \overline{0},$$

so that f is well-defined.

For the inverse mapping g, suppose $\overline{x} = \overline{y}$. Then $\overline{x - y} = \overline{0}$. So by Lemma 4.1, $x - y = pz$ for some $z \in V$. We have

$$px - py = p(x - y) = p^2 z = 0.$$

So g is well-defined. By inspection we see that both f and g are $\mathbb{Z}/p$-linear functions, and so the proof is complete. $\square$

The main proposition of this section shows that any lift of a basis of $\overline{V}$ is a basis of V. Such bases will be useful for constructing linear maps out of V. That is, if one defines any basis of V, then this map extends linearly to all of V.

Proposition 4.3. Let $\{e_i\}$ be a basis for $\overline{V}$. For each e_i, let $\tilde{e}_i$ in V be any lift of e_i. Then $\{\tilde{e}_i\}$ is a basis for V.

Proof. We first show that the set $\{\tilde{e}_i\}$ is linearly independent. Suppose $\alpha_i \in \mathbb{Z}/p^2$ with

$$\sum_i \alpha_i \tilde{e}_i = 0. \tag{2}$$

Projecting to $\overline{V}$ we obtain $\sum_i \overline{\alpha}_i e_i = \overline{0}$. Since $\{e_i\}$ is a basis for $\overline{V}$, we must have that $\overline{\alpha_i} = \overline{0}$ for every i. So each $\alpha_i = p\beta_i$ for some β_i in $\mathbb{Z}/p^2$. Thus, (2) gives that $\sum_i \beta_i \cdot p\tilde{e}_i = 0$ in pV. Under the isomorphism given in Lemma 4.2, we have $\sum_i \overline{\beta}_i e_i = \overline{0}$ in $\overline{V}$. Since the set $\{e_i\}$ is linearly independent, each $\overline{\beta_i} = \overline{0}$, so each $\beta_i = p\gamma_i$ for some γ_i in $\mathbb{Z}/p^2$. This gives that each $\alpha_i = p\beta_i = p^2\gamma_i = 0$. So the set $\{\tilde{e}_i\}$ is linearly independent.

We next show that $\{\tilde{e}_i\}$ spans V. Let $x \in V$. Since the set $\{e_i\}$ is a basis for $\overline{V}$, there are $\alpha_i \in \mathbb{Z}/p^2$ such that $\sum_i \overline{\alpha_i} e_i = \overline{x}$. So for some $y \in V$,

$$x = py + \sum_i \alpha_i \tilde{e}_i. \tag{3}$$

Under the isomorphism given in Lemma 4.2, the element py in pV is mapped to $\overline{y}$ in $\overline{V}$. Since the e_i form a basis for $\overline{V}$, there exist β_i in $\mathbb{Z}/p^2$ such that $\sum_i \overline{\beta_i} e_i = \overline{y}$. Thus $pz + \sum_i \beta_i \tilde{e}_i = y$ for some $z \in V$. Substituting this into (2) gives

$$x = p\left(pz + \sum \beta_i \tilde{e}_i\right) + \sum_i \alpha_i \tilde{e}_i.$$

Simplifying gives $x = \sum_i (\alpha_i - p\beta_i)\tilde{e}_i$, so that x is in the span of $\{\tilde{e}_i\}$, as desired. □

For the map ψ with domain $\overline{V}$ and target $\overline{W}$, recall that $\operatorname{coker}\psi$ is defined as the quotient $\overline{W}/\psi(\overline{V})$. Our next lemma shows that we may regard the Bockstein homomorphism as a map $\beta : \ker\psi \to \operatorname{coker}\psi$. The techniques used in the proof are similar to the techniques used in Lemma 4.2.

Lemma 4.4. The map f from $pW/\phi(pV)$ to $\operatorname{coker}\psi$ defined by

$$f : pw + \phi(pV) \to \overline{w} + \psi(\overline{V})$$

is an isomorphism.

Proof. We must show that f and its inverse mapping g are well-defined.

To show that f is well-defined, suppose $pw + \phi(pV) = pw' + \phi(pV)$ in $pW/\phi(pV)$. We must show that $\overline{w} - \overline{w'}$ is in $\psi(\overline{V})$. We have that $p(w - w') \in \phi(pV)$. Thus $p(w - w') = p\phi(v)$ for some $v \in V$. By Lemma 4.1, we have $w - w' - \phi(v) = py$. Thus $\overline{w} - \overline{w'} = \overline{\phi(v)} = \psi(\overline{v})$. So $\overline{w} - \overline{w'}$ is in $\psi(\overline{V})$, and this shows that f is well-defined.

We next want to show that the inverse mapping g is well-defined. Suppose that $\overline{w} + \psi(\overline{V}) = \overline{w'} + \psi(\overline{V})$. We must show that $pw + \phi(pV) = pw' + \phi(pV)$.Since $\overline{w} - \overline{w'} + \overline{\phi}(\overline{V}) = \overline{0} + \overline{\phi}(\overline{V})$, there exists a $\overline{v} \in \overline{V}$ with $\overline{w} - \overline{w'} = \overline{\phi}(\overline{v})$. Thus $w - w' - \phi(v) = px$ for some x, which, by Lemma 4.1 gives $p[w - w' - p\phi(v)] = 0$. So $pw + \phi(pV) = pw' + \phi(pV)$. By inspection, f and g are both linear, and the proof is complete. □

4.3 Spaces of Linear Maps

We should like to further investigate the connection between a map $\psi : \overline{V} \to \overline{W}$ and the Bockstein homomorphisms induced by a map $\phi : V \to W$ such that $\overline{\phi} = \psi$. For this section, we shall treat ψ as a fixed $\mathbb{Z}/p$-linear map from $\overline{V}$ to $\overline{W}$.

Definition 4.5. Let V and W be $\mathbb{Z}/p^2$-modules. Let $\overline{V}$ and $\overline{W}$ be the reductions of V and W modulo p. Let ψ be a fixed $\mathbb{Z}/p$-linear map from $\overline{V}$ to $\overline{W}$. Define L_ψ to be the collection of all maps from V to W whose reduction modulo p is ψ.

It will also be useful in this section to define a basis for $\overline{V}$, which will lift to a basis for V.

Definition 4.6. Let V, $\overline{V}$, and ψ be as in Definition 4.5. Let $\{e_i\} \cup \{f_j\}$ be a basis for $\overline{V}$ such that $\{e_i\}$ is a basis for the subspace $\ker \psi$ of $\overline{V}$. For each i, let $\tilde{e_i}$ in V be a lift of e_i. For each j let $\tilde{f_j}$ in V be a lift of f_j.

By Proposition 4.3, $\{\tilde{e_i}\} \cup \{\tilde{f_j}\}$ is a basis for V. If the map $\psi : \overline{V} \to \overline{W}$ is not the zero map, then we know that L_ψ is not a vector space, for in this case, 0 is not in L_ψ. This fact, along with the next lemma, gives that L_ψ is a vector space if and only if ψ is the zero map.

Lemma 4.7. The space L_0 with pointwise addition and $\mathbb{Z}/p$ scalar multiplication defined by

$$\overline{\alpha} \cdot \phi := \alpha \cdot \phi,$$

where α is in $\mathbb{Z}/p^2$ and ϕ is in L_0, is a $\mathbb{Z}/p$-vector space. In particular, if V has dimension n and W has dimension m, then L_0 is a $\mathbb{Z}/p$-vector space of dimension $m \cdot n$.

Proof. We shall only show that this scalar multiplication is well-defined, as the other parts of the proof are straightforward. Let α_1 and α_2 be in $\mathbb{Z}/p^2$ with $\overline{\alpha_1} = \overline{\alpha_2}$. Let ϕ be in L_0 and let $v \in V$. Then $\alpha_1 - \alpha_2 = p\beta$ for some β in $\mathbb{Z}/p^2$ and $\phi(v) = pw$ for some w in W, because $\overline{\phi} = 0$. So we have

$$\begin{aligned} \overline{\alpha_1} \cdot \phi(v) - \overline{\alpha_2} \cdot \phi(v) &= \alpha_1\phi(v) - \alpha_2\phi(v) \\ &= (\alpha_1 - \alpha_2)\phi(v) \\ &= (p\beta)(pw) \\ &= p^2\beta w \\ &= 0, \end{aligned}$$

which shows that this scalar multiplication in $\mathbb{Z}/p^2$ is well-defined. $\square$

Let ϕ_0 be any element of L_ψ. Then $\phi_0 + L_0 = L_\psi$, so we may regard L_ψ as a coset of L_0. It will be useful however to choose a particular $\phi_0 \in L_\psi$ whenever we wish to regard L_ψ as a coset of L_0. For this, we need only define ϕ_0 on the basis $\{\tilde{e_i}\} \cup \{\tilde{f_j}\}$ given in Definition 4.6.

Remark 4.8. When we regard L_ψ as a coset, we shall choose ϕ_0 such that $\phi_0(\tilde{e_i}) = 0$ for all i and $\phi_0(\tilde{f_j})$ be any value whose reduction modulo p is $\psi(f_j)$.

We are now ready to count the number of elements in L_ψ.

Lemma 4.9. For any $\psi : \overline{V} \to \overline{W}$, the set L_ψ has p^{mn} elements.

Proof. Lemma 4.7 tells us that L_0 is a $\mathbb{Z}/p$-vector space, but by definition, L_0 also is a $\mathbb{Z}/p^2$-submodule of $\mathrm{Hom}(V, W)$. When we regard L_ψ as $\phi_0 + L_\psi$, where ϕ_0 is as defined in Remark 4.8, this addition occurs in a $\mathbb{Z}/p^2$-submodule. So, while L_ψ is not a translate of L_0 as a $\mathbb{Z}/p$-vector space, we still know that L_ψ has the same number of elements as L_0. This information, along with Lemma 4.7, completes the proof. $\square$

4.4 The Bockstein Homomorphism

What follows is a short review of the Bockstein homomorphism in the context that is relevant for our study of cohomology operations with randomness. Several references cover the Bockstein homomorphism and cohomology operations in more generality. See, for example, [**?**].

As in Section 4.2, let V and W be $\mathbb{Z}/p^2$ free-modules with coefficients in $\mathbb{Z}/p^2$. We have the following short exact sequences:

$$0 \to pV \hookrightarrow V \twoheadrightarrow \overline{V} \to 0 \quad \text{and} \quad 0 \to pW \hookrightarrow W \twoheadrightarrow \overline{W} \to 0,$$

where $\overline{V}$ and $\overline{W}$ are the reductions of V and $W \mod p$.

Consider a $\mathbb{Z}/p^2$-linear map ϕ from V to W. Let ψ be the map from $\overline{V}$ to $\overline{W}$ induced by ϕ. Then the Snake Lemma [**?**] defines a map β with domain $\ker \psi$ and target $pW/\phi(pV)$. The following diagram illustrates the Snake Lemma.

$$\begin{tikzcd}
 & & \operatorname{Ker}\psi \arrow[d] \arrow[dddll, bend left, "\beta"'] \\
pV \arrow[r] \arrow[d] & V \arrow[r] \arrow[d, "\phi"] & \overline{V} \arrow[d, "\psi"] \\
pW \arrow[r] \arrow[d] & W \arrow[r] & \overline{W} \\
pW/\phi(pV) & &
\end{tikzcd}$$

More precisely, for $\overline{v} \in \ker\psi$, choose any representative $v \in V$ of $\overline{v}$. Since the squares in the above diagram commute, we have $\overline{\phi(v)} = \psi(\overline{v}) = \overline{0}$. So $\phi(v) = pw$ for some $w \in W$. Define the Bockstein homomorphism β from $\ker\psi$ to $pW/\phi(pV)$ by

$$\beta(\overline{v}) := pw + \phi(pV).$$

The following diagram shows the process described above.

$$\begin{tikzcd}
 & v \arrow[r, mapsto] \arrow[d, mapsto, "\phi"] & \overline{v} \\
pw \arrow[r, mapsto] \arrow[d] & pw = \phi(v) & \\
pw + \phi(pV) & &
\end{tikzcd}$$

By construction, the target of β is $pW/\phi(pV)$. However, by Lemma 4.4, we know that $pW/\phi(pV)$ is isomorphic to $\operatorname{coker}\psi$. So henceforth we shall regard β as a map into $\operatorname{coker}\psi$.

Remark 4.10. We note here that if one regards an arbitrary chain complex, the map β is often called a connecting homomorphism. When the chain complex is generated by a topological space, the map β is called the Bockstein homomorphism. If we regard ϕ as the map between V and W in the following chain complex

$$\cdots \longrightarrow 0 \longrightarrow V \xrightarrow{\phi} W \longrightarrow 0 \longrightarrow \cdots,$$

and consider the reduced chain complex

$$\cdots \longrightarrow 0 \longrightarrow \overline{V} \xrightarrow{\psi} \overline{W} \longrightarrow 0 \longrightarrow \cdots,$$

then the only possible non-trivial homology groups of this chain complex are $\ker\psi$ and $\operatorname{coker}\psi$. Although we are in a strictly algebraic setting, we shall continue to refer to the map β as the Bockstein homomorphism between $\ker\psi$ and $\operatorname{coker}\psi$.

Remark 4.11. The Bockstein homomorphism is often constructed in the case where V and W are $\mathbb{Z}$-modules. In this case, first reduce V and W to $\mathbb{Z}/p^2$ modules, and then apply the above construction.

In this section we have described how to every $\phi \in L_\psi$, there is a unique Bockstein homomorphism $\beta_\phi : \ker\psi \to \operatorname{coker}\psi$. This fact defines the following map.

Definition 4.12. Define Γ to be the map from L_ψ to $\operatorname{Hom}(\ker\psi, \operatorname{coker}\psi)$ that sends ϕ in L_ψ to the unique Bockstein homomorphism β_ϕ in $\operatorname{Hom}(\ker\psi, \operatorname{coker}\psi)$ that is given by ϕ.

Composing Γ with addition by ϕ_0 gives a well defined set map B with domain L_0 and target

$$\operatorname{Hom}(\ker\psi, \operatorname{coker}\psi).$$

This is shown in the following diagram.

$$\begin{tikzcd} L_0 \arrow[r, "+\phi_0"] \arrow[rr, bend left, "B"] & L_\psi \arrow[r, "\Gamma"] & \operatorname{Hom}(\ker\psi, \operatorname{coker}\psi) \end{tikzcd}$$

We should like to examine the properties of this map. The map from L_0 to L_ψ given by adding ϕ_0 is a bijection. The next lemma shows that the map Γ is onto, which shows that B is also onto. In particular, every $\mathbb{Z}/p$ linear map from the kernel of ψ to the cokernel of ψ is the Bockstein homomorphism of some $\phi : V \to W$ that induces ψ.

Lemma 4.13. The map Γ from L_ψ to $\operatorname{Hom}(\ker\psi, \operatorname{coker}\psi)$ is onto.

Proof. Let $\beta \in \operatorname{Hom}(\ker\psi, \operatorname{coker}\psi)$. We must find a $\phi \in L_\psi$ such that the Bockstein homomorphism of ϕ is β. Let $\{e_i\} \cup \{f_j\}$ and $\{\tilde{e_i}\} \cup \{\tilde{f_j}\}$ be bases of $\overline{V}$ and V as defined in Definition 4.6. We shall define ϕ on the basis for V and then extend linearly to define ϕ on all of V. We must then show that the Bockstein homomorphism β_ϕ of ϕ is equal to β.

For each i, we know that e_i is in the domain of β. So $\beta(e_i) = \overline{w_i} + \psi(\overline{V})$ for some w_i in W. Define $\phi(\tilde{e_i}) = pw_i$. Define $\phi(\tilde{f_j})$ to be any value in W whose reduction modulo p is $\psi(f_j)$. Then $\overline{\phi(e_i)} = \overline{pw_i} = 0 = \psi(e_i)$ and $\overline{\phi(\tilde{f_j})} = \psi(f_j)$. This shows that $\overline{\phi} = \psi$. In particular ϕ is in L_ψ.

By construction, $\beta_\phi(e_i) = \overline{w_i} + \psi(\overline{V}) = \beta(e_i)$. Since β_ϕ is equal to β on the basis of $\ker\psi$, they are equal as $\mathbb{Z}/p$-linear functions. $\square$

Lemma 4.14. The map B is a $\mathbb{Z}/p$-linear map.

Proof. Let ϕ and ϕ' be in L_0. We must show that $\Gamma(\phi+\phi'+\phi_0) = \Gamma(\phi+\phi_0)+\Gamma(\phi'+\phi_0)$, for $\phi_0 \in L_\psi$ described in Remark 4.8. So, for a basis $\{e_i\}$ of $\ker\psi$, it suffices to show that

$$\Gamma(\phi+\phi'+\phi_0)(e_i) = \Gamma(\phi+\phi_0)(e_i)+\Gamma(\phi'+\pi_0)(e_i)$$

Let $\tilde{e_i}$ be any lift of e_i. Then $\phi_0(\tilde{e_i}) = 0$ by construction. Also,

$$\overline{(\phi+\phi')(\tilde{e_i})} = \overline{\phi(\tilde{e_i})} + \overline{\phi'(\tilde{e_i})} = 0, \tag{4}$$

and

$$\overline{\phi(\tilde{e_i})} = \overline{\phi'(\tilde{e_i})} = 0, \tag{5}$$

because ϕ and ϕ' are in L_0. Thus there are w_i and w_i' in W with $\phi(\tilde{e_i}) = pw$ and $\phi'(\tilde{e_i}) = pw'$. Thus by (5) we know that

$$\Gamma(\phi+\phi_0)(e_i)+\Gamma(\phi'+\phi_0)(e_i) = \left(\overline{w_i}+\psi(\overline{V})\right) + \left(\overline{w_i'}+\psi(\overline{V})\right) = \overline{w_i}+\overline{w_i'}+\psi(\overline{V}).$$

Equations (4) and (5) together give that

$$\Gamma(\phi+\phi'+\phi_0)(e_i) = \overline{w_i}+\overline{w_i'}+\psi(\overline{V}).$$

So $\Gamma(\phi+\phi'+\phi_0)(e_i) = \Gamma(\phi+\phi_0)(e_i)+\Gamma(\phi'+\phi_0)(e_i)$, as desired. $\square$

4.5 Counting

Let V and W be $\mathbb{Z}/p^2$-modules of dimensions n and m respectively. In Section 4.3 we defined L_ψ as the collection of all maps from V to W whose reduction modulo p is ψ. We then found that L_ψ has $p^{m\cdot n}$ elements. We should next like to answer the following question: Given a Bockstein homomorphism β, which is in $\hom(\ker\psi\,\mathrm{coker}\,\psi)$, how many ϕ in L_ψ have β as their Bockstein homomorphisms? To answer this question, we shall first look at the size of $\Gamma^{-1}(\beta)$.

Lemma 4.15. Let $k := \dim(\ker\psi)$. Then the space $\Gamma^{-1}(\beta)$ has $p^{(m+k)(n-k)}$ elements.

Proof. Since translation by the ϕ_0 given in Remark 4.8 is a bijection, we know that $B^{-1}(\beta)$ has the same size as $\Gamma^{-1}(\beta)$. Since B is a linear map, we also know that $B^{-1}(0)$ has the same size as $B^{-1}(\beta)$. Thus $\Gamma^{-1}(\beta)$ has the same size as $B^{-1}(0)$, so we shall find the size of $B^{-1}(0)$.

Since V and W have dimension n and m respectively, we know by Proposition 4.3 that $\overline{V}$ and $\overline{W}$ also have dimensions n and m respectively. Recall that $\operatorname{coker}\psi$ is defined as $\overline{W}/\psi(\overline{V})$. Let $k := \ker(\psi)$. Since ψ is a $\mathbb{Z}/p$-linear map, by the Rank-Nullity Theorem, we have that $n = k + \dim(\psi(\overline{V}))$. So $\dim(\psi(\overline{V})) = n - k$. Thus $\dim(\operatorname{coker}(\psi)) = m - (n-k)$. From this, we have that the number of elements in $B^{-1}(0)$ is

$$p^{mn-k(m-n+k)} = p^{(m+k)(n-k)}.$$

□

We now come to our main result. Recall that a choice of random function from V to W is the same as choosing a random m by n matrix.

Theorem 4.16. Let ϕ be an m by n matrix whose entries are chosen i.i.d. from the discrete uniform distribution on $\{0, 1, 2, \ldots, p^2 - 1\}$. Let $\psi = \overline{\phi}$. Let β_ϕ be the Bockstein homomorphism defined by ϕ as in Definition 4.12. Note that β_ϕ is a random variable. Let β be in $\hom(\ker\psi, \operatorname{coker}\psi$. Then

$$\mathbb{P}\big(\beta_\phi = \beta \big| \overline{\phi} = \psi\big) = \frac{1}{p^{k(m-n+k)}}.$$

Proof. We know from Remark 4.9 that L_ψ has p^{mn} elements. By Lemma 4.13, we know that Γ is onto, and by Lemma 4.15, we know that the size of $\hom(\ker\psi, \operatorname{coker}\psi)$ is $p^{mn-k(m-n+k)}$. □

REFERENCES

[1] Pierre Arnoux and Alexis Marin, *The kühnel triangulation of the complex projective plane from the view point of complex crystallography part ii*, Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics **45** (1991), no. 2, 167–244.

[2] Michael Francis Atiyah and Ian Grant Macdonald, *Introduction to commutative algebra*, vol. 2, Addison-Wesley Reading, 1969.

[3] Andrew J Blumberg, Itamar Gal, Michael A Mandell, and Matthew Pancia, *Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces*, Foundations of Computational Mathematics **14** (2014), no. 4, 745–789.

[4] Béla Bollobás, *Random graphs*, Springer, 1998.

[5] Frédéric Chazal, Brittany T Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman, *Robust topological inference: Distance to a measure and kernel distance*, arXiv preprint arXiv:1412.7197 (2014).

[6] Vin De Silva and Robert Ghrist, *Coverage in sensor networks via persistent homology*, Algebraic & Geometric Topology **7** (2007), no. 1, 339–358.

[7] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, Aarti Singh, et al., *Confidence sets for persistence diagrams*, The Annals of Statistics **42** (2014), no. 6, 2301–2339.

[8] Allen Hatcher, *Algebraic topology*, Cambridge University Press, 2002.

[9] Matthew Kahle, *Topology of random clique complexes*, Discrete Mathematics **309** (2009), no. 6, 1658–1671.

[10] ———, *Random geometric complexes*, Discrete & Computational Geometry **45** (2011), no. 3, 553–573.

[11] Ann B Lee, Kim S Pedersen, and David Mumford, *The nonlinear statistics of high-contrast patches in natural images*, International Journal of Computer Vision **54** (2003), no. 1-3, 83–103.

[12] PY Lum, G Singh, A Lehman, T Ishkanov, Mikael Vejdemo-Johansson, M Alagappan, J Carlsson, and G Carlsson, *Extracting insights from the shape of complex data using topology*, Scientific reports **3** (2013).

[13] Robert E Mosher and Martin C. Tangora, *Cohomology operations and applications in homotopy theory*, Dover Publications, 2008.

[14] Monica Nicolau, Arnold J Levine, and Gunnar Carlsson, *Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival*, Proceedings of the National Academy of Sciences **108** (2011), no. 17, 7265–7270.

[15] Mathew Penrose, *Random geometric graphs*, vol. 5, Oxford University Press Oxford, 2003.

[16] Gurjeet Singh, Facundo Mémoli, and Gunnar E Carlsson, *Topological methods for the analysis of high dimensional data sets and 3d object recognition.*, SPBG, Citeseer, 2007, pp. 91–100.

ABSTRACT

COHOMOLOGY OPERATIONS ON RANDOM SPACES

by

MATTHEW ZABKA

May 2016

Advisor: Dr. Daniel Isaksen

Major: Mathematics

Degree: Doctor of Philosophy

Topology has recently received more attention from statisticians as some its tools have been applied to understanding the shape of data. In particular, a data set can generate a topological space, and this space's topological structure can give us insight into some properties of the data. This framework has made it necessary to study random spaces generated by data. For example, without an understanding of the probabilistic properties of random spaces, one cannot conclude with any degree of confidence what the tools of topology tell us about a data set. While some results are known about the cohomological structure of a random space, not much is known about how cohomology operations behave on random spaces. This dissertation proves some results about the asymptotic properties of cohomology operations on random spaces and discusses the idea of a random Bockstein operation in a related purely algebraic context.

AUTOBIOGRAPHICAL STATEMENT

MATTHEW ZABKA

Education :

- Ph.D. in Mathematics, May 2016, Wayne State University
- M.A. in Statistics, May 2012, University of Michigan
- M.A. in Mathematics, May 2010, Wayne State University
- B.A. in Mathematics, Economics, and German May 2007, Saint Cloud State University

Awards

- Dr. Chorng-Shi Houh Award, April 2016, Wayne State University
- Outstanding Teaching Award, April 2012, University of Michigan
- Munich Graduate Exchange Fellowship, 2009-2010, Wayne State University/Ludwig-Maximillian University
- Mathematics Scholarship, November 2006, Saint Cloud State University